Will Epperson

I'm a Ph.D. student in the HCII at CMU advised by Dominik Moritz and Adam Perer.

My research interests lie in **developing interactive data science tools for both experts and non-experts**. I am interested in how we can help analysts understand their data through interactive visualization, recommended analysis, and models. My commonly used research methods include building stand-alone systems for data analysis, building extensions to existing tools like Jupyter, and running human studies experiments to evaluate these systems.

willepperson.com
willepp@cmu.edu
CV PDF
@w_epperson
@willeppy
Google Scholar
LinkedIn

Education

August 2020 - Present

Ph.D. in Human Computer Interaction
Carnegie Mellon University
Advisors: Dominik Moritz, Adam Perer
Sample Coursework: HCI Process and Theory, Computational Medicine, Human Judgement and Decision Making

August 2016 — May 2020

B.S. in Computer Science
Georgia Institute of Technology
GPA: 4.0, Summa Cum Laude, threads in Intelligence and Modeling/Simulation
Sample Coursework: Machine Learning, Deep Learning, Computer Vision, Computer Architecture, Algorithms, Computer Simulation, Information Visualization

Publications

Leveraging Analysis History for Improved In Situ Visualization Recommendation
Will Epperson, Doris Jung-Lin Lee, Leijie Wang, Kunal Agarwal, Aditya Parameswaran, Dominik Moritz, Adam Perer
Solas is a visualization recommendation tool that uses the history of analysis for in situ recommendations in Jupyter.
EuroVis 22: Eurographics Conference on Visualization (EuroVis). Rome, Italy, 2022.
Project PDF Code BibTeX

Strategies for Reuse and Sharing among Data Scientists in Software Teams
Will Epperson, April Yi Wang, Robert DeLine, Steven M. Drucker
Interviews and a survey with 149 data scientists at Microsoft revealed five distinct strategies for sharing and reusing analysis code along with factors that encourage or discourage reuse.
ICSE 22: ACM International Conference on Software Engineering (ICSE). Pittsburgh, PA, 2022.
Project PDF Recording BibTeX

Diff in the Loop: Supporting Data Comparison in Exploratory Data Analysis
April Yi Wang, Will Epperson, Robert DeLine, Steven M. Drucker
Diff in the Loop supports tracking, comparing, and visualizing differences in datasets during iterative data analysis.
SIGCHI 22: ACM Symposium on Computer Human Interaction (CHI). New Orleans, LA, 2022.
Project PDF BibTeX

RECAST: Interactive Auditing of Automatic Toxicity Detection Models
Austin P. Wright, Omar Shaikh, Haekyu Park, Will Epperson, Muhammed Ahmed, Stephane Pinel, Diyi Yang, Duen Horng (Polo) Chau
Interactive Auditing of Automatic Toxicity Detection Models
24th ACM Conference on Computer-Supported Cooperative Work & Social Computing. 2021.
Project PDF BibTeX

FairVis: Visual Analytics for Discovering Intersectional Bias in Machine Learning
Angel Cabrera, Will Epperson, Fred Hohman, Minsuk Kahng, Jamie Morgenstern, Duen Horng (Polo) Chau
Discovering intersectional ML Bias through interactive visualization.
IEEE Conference on Visual Analytics Science and Technology (VAST). Vancouver,

Canada, 2019.

Project Demo PDF Blog Recording Slides Code BibTeX

Talks

FairVis

May 2019

VIS 19: IEEE Visualization Conference

Strategies for Reuse and Sharing among Data Scientists in Software Teams

May 2022

ICSE 22: ACM International Conference on Software Engineering

Honors and Awards

2019

PURA: President's Undergraduate Research Award

$1500 research grant to continue work on FairVis project

2016

Stamps President's Scholarship

Full ride scholarship given to 40 incoming freshman at Georgia Tech

Research Experience

August 2020 - Present

Carnegie Mellon University, Pittsburgh, PA

Graduate Researcher, Data Interaction Group (DIG)

Advisor: Dominik Moritz, Adam Perer

Member of the DIG research group, working on novel data visualizations, ML interpretation techniques, and interactive data systems.

Relevant Skills: Python, Javascript

January 2019 - May 2020

Georgia Institute of Technology, Atlanta, GA

Undergraduate Researcher, Polo Club of Data Science

Advisor: Duen Horng (Polo) Chau

Member of the Polo Club of Data Science working on novel data visualizations to find fairness issues in Machine Learning models

Relevant Skills: Python, Javascript

January 2018 - May 2019

Georgia Institute of Technology, Atlanta, GA

Undergraduate Researcher, Automated Algorithm Design

Advisor: Jason Zutty, Greg Rohling

Worked on EMADE algorithm design engine to implement sentiment analysis pipeline to analyze news articles to aid in predicting stock price movements using genetic algorithms. Led project to visualize the genetic algorithm evolution process.

Relevant Skills: Python, Javascript

Industry Experience

Summer 2021

Microsoft Research, Redmond, WA

Research Intern, VIDA Group

Mentor: Steve Drucker, Rob DeLine

Research intern working on data science tools. Lead project around reuse and sharing in data science, in submission to ICSE 2022. Also involved with project around visualizing data frame differenes in submission to CHI 2022.

Relevant Skills: Python, Typescript

Summer 2019

Point72 Asset Management, New York, NY

Data Analytics Intern, Market Intelligence Group

Mentor: Trevor Rempel

Worked as Data Scientist in alternative data space to clean, model, and understand large datasets

Relevant Skills: Python, Distributed Computing in Spark

Summer 2018

Ultimate Software, Weston, FL

Software Development Intern, Innovation Strategies Team

Mentor: Joseph Cutrono

Designed and developed Slack app to integrate with the UltiPro HR management tool. App published to Slack app store.

Relevant Skills: Typescript, REST API development

Summer 2015

The Home Depot, Atlanta, GA
Software Development Intern
Developed web app for tracking candidate progress throughout hiring process for internal HR use.
Relevant Skills: Java, HTML/CSS/Javascript

Mentees

During my PhD, I have had the pleasure of mentoring the following undergraduate students on research projects.

Summer 2021 - Fall 2021

Leijie Wang
Visualization recommendation for python in notebooks using history

Fall 2021 - Spring 2022

Asad Sheikh
Visualization recommendation for SQL using history

Spring 2022+

Vaishnavi Gorantla
Fact generation from data and presentation as text

Teaching

August 2017 - December 2018

Undergraduate Teaching Assistant
Georgia Institute of Technology, Atlanta, GA
Intro to Database Systems (CS 4400), Instructor: Monica Sweat
Designed projects, held office hours and graded for relational databases class.

Leadership & Activities

January 2019 - May 2020

Student Ambassador
Georgia Institute of Technology Alumni Association
Serve as official representative of the Institute at events/tours for alumni, prospective students, and special guests.

January 2018 - May 2019

Executive Board Member -- Threads Co-chair
Stamps Scholars National Convention 2019
Executive board member of Stamps Scholars National Convention, a 3-day conference with over 700 student attendees. Responsible for 20-person committee that plans and coordinates the different content threads of the convention.

Sample Projects

December 2018

ICLR'19 Reproducibility Challenge
Implemented the architecture and reproduced results for an ICLR'19 submission using GANs to de-correlate sensitive data, titled Generative Adversarial Models for Learning Private and Fair Representations.

January 2018 - December 2018

Atlanta Crime Map
Led a team as part of Data Science Club at GT to analyze and visualize crime data on and around GT's campus to provide insight into crime frequency and details for GTPD.

February 2018

Citadel Datathon
Created supervised learning system to predict dangerous road areas to help make cities safer at Citadel and Correlation One's Datathon at Georgia Tech. Team placed top 3.

February 2017 - December 2017

RECONSO Research
Developed interface between core state machine and battery systems on Avionics team of student-led cube satellite project. Learned C while working on project.

Skills

Programing Languages: Python (Advanced), Javascript/Typescript (Intermediate), HTML (Intermediate), SQL (Intermediate), Java (Intermediate), C (Basic)
Toolkits, Frameworks, Software: Pytorch, Scikit-learn, Git, VegaLite, D3, Tableau, MacOS, Windows, Linux
Natural Languages: English (Native), Spanish (Advanced)